STATEWIDE CLIENT INDEX USER GROUP MEETING
APRIL 17, 2001 MINUTES

Attendees:

Ba, Ho Le	CalWIN
Baumgartner, Eric	EDS-HFP
Bell, Lesley	DSS Fraud
Bornstein, Marty	DHS ITSD
Bowen, Christine	C-IV
Brill, Roger	DHS ITSD
Butts, Diana	Riverside County
Cisneros, Shirlee	DHS ITSD
Clendenin, Cindy	LEADER
Colate, Jennifer	CDSS-Fraud
Enrico, Larry	C-IV
Hamilton, Tana	ISAWS
Hanna, Waltina	HHSDC-SFIS
Hartz, Barbara	DHS MCH
Kha, Ka	ISAWS
Lara, Elena	DHS MEB
Murphy, David	C-IV
Olson, Pete	DHS ITSD
Parman, Debbie	Nevada County
Perry, Sharon	DHS ITSD
Pierce, Rick	Stanislaus County
Ragazzi, Ginger	LEADER
Richman, Seth	C-IV
Rundall, Julie	DHS CMS
Sakauye, Dave	HHSDC-SFIS
Sabel, Kathy	DHS ITSD
Sheldon, Meg	CWDA
Swedlow, Jim	LEADER
Vellanoweth, Angelica	HHSDC-SFIS
Wang, Jina	DHS ITSD

Purpose of the meeting

The purpose of this meeting was to discuss the potential system changes to the Statewide Client Index (SCI) application that will make SCI more useful and provide more accurate data to our customers.

Changes to the SCI Query process

The SCI application assists public health and public assistance applications in linking client data across programs and across time. As DHS staff gain more familiarity with the data and with our customer's process requirements, we'll refine our processing to provide a better set of data. The heart of SCI is the query process to identify whether a client is already known on the file or is a new client who needs to be assigned a new unique ID, called a Client Index Number (CIN).

SCI has both a batch match and an online inquiry process.
There are some basic differences between the online interactive query process and the batch process. Most SCI customers query SCI interactively. These users typically have the option of getting more information in order to make a decision if the client is already known. The client may be present or could be contacted. Users could check data on other systems, most notably MEDS, but also check the photo if the client is known to SFIS. For these interactive queries, SCI's goal is to provide a list of potential matches and let the customer make the decision.

There are three situations where CINs are or were assigned in batch:

- MEDS does a standard nightly update process:

- Those Los Angeles county districts, which had not yet implemented LEADER, assigned CINs to their clients in batch;

- Typically when customers initially implement SCI, DHS does a file load process to provide CINs for existing clients. CalWIN staff plans their conversion to use CINs next year.

In the batch realm, the computer must make the decision as whether the client is known or not. If the computer decides that the client is known to the system, when he or she really is not known, it could adversely affect the client's ability to receive services from Medi-Cal. A new ID card, often called a BIC card, which is required by providers as proof of eligibility before services are provided, may not be issued. Another potential consequence of a bad decision is that eligibility could be linked to the wrong record, so when a provider does a card swipe to check eligibility, it could show as not eligible when the client really is eligible.

Since the batch process does a "yes" or "no" decision whether the client is known or not and there are consequences of making a wrong decision, the rules for a batch match are relatively strict. Since the interactive process does not make a "yes" or "no" decision, and there is an assumption that the user will be making an informed decision, SCI can return multiple potential matches.

The other distinction is that SCI has a different process for doing a key-based search than a name search. The typical key-based search is an SSN, but other keys that could be used are the Client Index Numbers and Alien Numbers.

In the batch process, when SCI performs a key-based search, it will check to see that the name and birthdate are reasonable matches. If so, SCI considers this to be a valid match. If not, SCI will attempt to perform a name-based search for this client. In the interactive key-based search, SCI returns any record that matches the input key regardless of differences between the input record and the SCI record in the birthdate or name fields. These processes are unchanged. What has changed and is changing is the interactive name query function.

Until February 2001, the process flow for the interactive name searches was as follows:

a) First, convert the input name to a phonetic name key

b) Then, select all records from SCI where the phonetic name key matches the key on the input record, and the gender and the birthyear both also match

c) Score the potential matches (for common names, there could be hundreds of potential matches) and return up to 25 records in order of the score.

The new process retains the conversion of the input name to a phonetic name key and selects all records from SCI where the phonetic name key matches the key on the input record, and the gender matches. However there were several changes in the process:

a) SCI now also selects those records where the birth year is plus or minus one year.

b) There is a minimum threshold score established to return potential matches, instead of returning all (or up to 25) matches, it will only return likely matches.

c) Logic is being added so that records that had matching middle initials were given more weight than those where either the input record or the SCI record had spaces in the middle name field.

d) Data cleansing logic has been developed to identify middle initials embedded in the first name and Appellations embedded in the last name, whose presence had been causing SCI to assign incorrect phonetic name keys.

The primary driving force for implementing these changes is to improve data quality and data integrity. But there are some other benefits of returning fewer potential matches. By pre-screening unlikely matches, there is less data for workers to evaluate. This will slightly improve response time and also reduce data communications costs. It allows the possibility of including additional data fields on the response without increasing the overall maximum record length.

SCI/MEDS Query Planned Changes

One of the SCI functions is to provide interactive queries of MEDS client data. LEADER and SFIS now utilize this function. The CalWIN consortium of counties also plans to use this function and has requested that this query, which is referred to as the OM transaction, include a few more data elements. (MEDS Name, OHC Source code, SSA DOB and DOB Validation code)

The target date for DHS to complete the programming is September 2001. Our plan is to put all new data elements at the end of the existing data flow, so that the current users of the OM trans do not need to change their programming.

There is a window of opportunity for current users of this transaction or possible future users to identify whether there are other MEDS data elements to be included.

The state public health applications, which access SCI via the CATS hub, have somewhat similar functionality available with the CATS 2100 transaction.

FEEDBACK FROM THE GROUP

Even though ISAWS does not currently utilize the OM transaction, they want to be informed about any changes to this query in the event that they do choose to use this function in the future.

Adding the MEDS Status Code to SCI

Unfortunately there are numerous clients with more than one record on SCI. The potential advantage of adding the MEDS status code on the query responses is to assist users to select a CIN in those cases when the client does appear to have more than one record on SCI. Many customers already have access to MEDS and utilize the MEDS address and family composition information to make help these decisions. If DHS were to add this enhancement, we would still encourage those customers to use MEDS as much as they do now.

The addition of this field would seemingly benefit only customers who either don't have direct access to MEDS or do have access to MEDS, but don't use it for file clearance.

In the proposed enhancement, MEDS would provide SCI with daily updates with eligibility information. There would also be a month end process to get in sync with MEDS as the month of eligibility changes. We are considering two alternatives:

- The simplest would be either Y (currently active) or N (not currently active).
- The alternative solution would have three codes -- currently active, recently active, or not recently active.

FEEDBACK FROM THE GROUP

It can definitely be useful to know the MEDS status, particularly for handling common names, but since the SCI OM transaction, which does a real time read of MEDS, already provides that information, then perhaps a simpler solution than changing the SCI query response would be for more SCI users to utilize the OM data in their process.

If we do incorporate the MEDS status into the query response in the future, then the consensus of the group would be to have a one-digit code that indicates approximately how long it has been since the client had been active.

Adding an Address-based matching to SCI

Incorporation of address matching within SCI is to include address as one of the variables that is scored as opposed to entering an address and see if any clients live at that address.

The purpose would be to improve the accuracy of handling common names when an SSN or other unique key is not available. It has been our experience that there are more problems with the handling of common names than less common names.

However, there are some concerns and constraints, which might not make this potential enhancement the best use of everyone's resources.

Potential Disadvantages:

1. Programming changes would be needed for both SCI and any application, which wanted to use address-based matching

2. SCI would need to do this in a way that would minimize the possibility of inappropriate disclosure of confidential information

3. Among the populations that we serve, a sizable percentage of our clients are very mobile, which could impact the accuracy of address match

4. Addresses can be obtained for most clients via MEDS, but for the approximately 8-9% of the SCI clients not known to MEDS, SCI would either not have an address or we would need to develop a mechanism to collect the address in SCI.

Potential solutions

1. Collect several occurrences of the client's 5-digit zip code. Match the current and prior zip codes in a similar manner to how SCI now uses the current and prior county code data.

It does not seem that providing the zip code is nearly as much of a confidentiality concern as if we were to provide the street address. Having zip codes on the file might be of assistance later if there was ever a GIS use of aggregate SCI data for research purposes. This might be a bigger help for users in large counties, than in small counties. In small counties, a match on the county code is meaningful, but much less so in large counties.

2. Collect street address data, but do not display it. Compare the input address data to the data on file in order to determine whether to adjust the match score up or down.

Whether we use the full street address or just the zip code, more points would be added for a match than subtracted for a mismatch.

FEEDBACK FROM THE GROUP

Address checking may be useful in those cases where the client is applying for services or benefits from multiple programs within the same relative timeframes, but may be counterproductive in many cases where the client had been known previously, but is currently in inactive status.

The concern expressed by the group is that this population can be so mobile that the old address on file can frequently not match the current address. There was also a concern about differences between mailing address and the residence address. Because of these factors, the group felt that if address matches are incorporated into SCI, then a mismatched address should not cause the matching score to be decreased, while a matching address should result in a sizable increase in the score.

It was recommended that, if we were to use address, that we should build this to use zip +4 or perhaps zip + 4 + 2.

There was a recommendation that the concept of incorporating address into the automated system should be manually field-tested. Counties would check the MEDS address to ascertain how frequently this agreed or disagreed with the address on the application. Based on their findings, the group could then make a better recommendation about
a) Whether we should include address as a variable to be scored by SCI and
b) If so, how much weight should be assigned.

Another possibility is to incorporate address matching into the batch process in order to minimize the issuance of pseudonumbers and duplicate CINs.

Utilizing CDL/ID as a Search Key on SCI

The legislature and the Administration have requested that DMV take steps to reduce the instances of identify theft. DMV has decided that part of this solution is to validate that SSNs reported actually belong to the client when folks apply for a new license or renew their license.

The Social Security Administration provides data to many state and local agencies. SSA's preference is to establish an agency in each state as the primary contact. In California, this is DHS. DMV contacted DHS when they wanted to establish a procedure to validate new SSNs and the backlog of over 20 million SSNs on file. As part of the package, for which DHS provided assistance, DMV agreed to allow us to collect CDL/IDs to be used in the administration of public assistance and public health applications.

DHS has collected DMV data in offline files for approximately 5 months. During that time, DHS has processed over 5 million DMV records and identified 1.3 Million CDL/IDs for clients on SCI. It was initially expected that it would take approximately 6-8 months for DMV to complete processing of their backlog. At that time, we anticipated having CDL/IDs for 5-6 million clients on SCI. However, things did not go as planned.

The number of records that do not validate is higher than DMV anticipated. Since a non-validated SSN requires follow-up, these result in workload for DMV staff. Thus, DMV is proceeding more slowly until they can reduce the non-validation rate. So now it may take a year or two for DMV to process the backlog, which obviously affects the number of CDL/IDs that we could capture on SCI, at least in the near future.

DHS will not post this data onto the SCI tables unless our customers feel that this would provide a benefit to your programs. Since this collection of Drivers License information is a byproduct of SSN Validation, then these clients who have CDL/IDs also have SSNs and in many cases, if the client applied for a public assistance program, then these SSNs have already been reported to the applications. So perhaps the use of CDL/ID as a search key is limited:

a) It could help validate that this is your client when the SSN has been transposed.

b) It could help validate that this is your client when the SSN on SCI matches your records, but the name or birthdate on SCI is different than your records.

c) It could be of assistance to those public health programs, which serve adults, where the SSN is not collected.
NOTE: only a relatively small number of the records, for which there is a CDL/ID, are associated with children.

Thus, the use of the DL in SCI does not seem nearly as promising to DHS staff as when we first found out that we could obtain the CDL/IDs.

FEEDBACK FROM THE GROUP

Adding the CDL/ID data may have some benefit for handling common names, but it is likely that other enhancements will have more benefit.

Utilizing the DMV Middle Name

Even if the use of the CDL/ID was a search key is not high on users' list of potential enhancements for SCI, DHS staff thought that there is another piece of data on the DMV files, which we had hoped, could be very useful. Currently around 30% of the SCI records have a Middle Initial and roughly 1.5% of our records contain middle names. Our thinking was that if we could increase these counts, this could help the processing for common names.

When clients apply for a driver's license, they typically include their middle name on the application. As DHS loaded CDL/ID data to the offline files, we saw that about 83% of the records had some data in the middle name field. About 5% of these records just had a middle initial; the remainder had middle name data.

This might be a vehicle to add millions of middle initials/names to SCI. However, DMV is trying to take steps to reduce the number of SSNs that do not validate. At the end of March, they stopped sending the middle names to SSA in the hope that this would help reduce the amount of exception processing that they needed to do. Over the past several weeks, omitting the middle name does not appear to impact the SSN validation rate, one way or another. If DMV uncovers another potential change which does reduce the amount of exception processing, then DHS will explore the possibility that DMV could renew sending middle names, but it's unlikely that DMV will modify their programming unless there is some benefit to their program.

If DHS updates SCI with the middle name data that we have received so far, we will have more complete middle name data than we do now, but not nearly as much as we had hoped. We could add middle initials to some records that now have spaces in this field. DHS could update some SCI records that have middle initials with a more complete middle name.

The question to be resolved is that with the decreased number of records that could be updated with middle name data is worthwhile, considering some of the issues that would need to be resolved. These include the following:

a) Would the addition of this data cause any problems for our customers?

If, for example, the only record on the SCI file for Marty Bornstein was from LEADER and if SCI updated the SCI record with a middle name of Steven, would it be misleading for a user in LA or another SCI user to see this record and assume that the LEADER knows that client by the full name?

FEEDBACK FROM THE GROUP

Adding middle name data to existing records was not perceived as a problem

b) Would we send a record of the update to the customers that know this client?

SCI creates a batch record, called a referral record, when two or more applications know a client and one of them changes some piece of data. The most common of these changes is to the county code. Any SCI customer could opt to obtain referrals, but so far, only ISAWS has chosen to do so.

One issue to be resolved is how SCI will generate the referrals. Referrals are a byproduct of the UPDATE process. The SCI UPDATE process will only allow updates from an application that "owns" data on SCI for this client. An application has "ownership" rights to a client record if it sends an add or a link transaction. DMV has not sent an add or link for this client. So, DHS would need to tweak our system to create referrals for DMV name updates.

FEEDBACK FROM THE GROUP

ISAWS definitely would want to be notified if we updated name data for their clients.

c) Another issue is DMV only allocated 7 bytes for the middle name,
So, if the middle name field contains all 7 bytes, it is not obvious whether the DMV Middle name field contains the full middle name or a partial middle name. Customers need to provide feedback about whether having an incomplete middle name is a problem.

If so, DHS could build some logic to handle some common names. For example, ROBERTO or LINDSEY, then move all 7 characters to the Middle name field; if CHRISTO, move CHRISTOPHER), but we obviously would not get all names. Another possibility would be to move the DMV Middle Initial if the DMV first name = the SCI first name; the DMV last name = the SCI last name; the SCI middle name contains spaces, and the DMV middle name contained either 1 or 7 characters.

If the SCI First/Last name matched the DMV first/last name, the SCI MN = spaces, and the DMV MN = 2 to 6 characters, then move the DMV MN to the SCI name.

On the other hand if having a potentially incomplete Middle name were not a problem, then our logic would be simpler.

FEEDBACK FROM THE GROUP

It is not a major problem to have truncated middle names as long as the applications are aware of this constraint in advance.

The group requested that DHS research how many records would actually
be updated with Middle Name data because there was a record on SCI with the same First and last names, but less data in the middle name than on
the DMV record.

Changing the Way that MEDS Collects CINs

Currently counties send an add transaction to MEDS and either the SSN or the County ID on the transaction matches a record on MEDS, then the current MEDS logic does not even look at the input CIN, but instead uses the CIN by which MEDS knows this client. This results in the application and MEDS being out of sync for that client. There are a couple of ramifications of this:

- Linking M/C children with HFP adults when SPE sends in a file of CINs to match for the HF Expansion program may not work as well as expected.

- When the customer enters the CIN by which he or she knows the client on a POS query, it is unlikely that the customer will find the right MEDS record.

- It might make it more difficult for HFP, CMS, ISAWS, and LEADER to track/utilize worker alerts.

- It could make it more difficult for workers to query MEDS.

This may not be a complete list of the potential problems, but it's enough to trigger a discussion for potential solutions.

The plan was to modify MEDS when a CIN is present on the add transaction. If this CIN differed from the CIN on MEDS, then MEDS would capture the input CIN on its CIN cross-reference file and make better use of this CIN to help reduce the types of problems listed above. Some items of interest:

A) We plan to begin collecting these other CINs from the time that we install this change.

B) If there is a reconciliation process between your application and SCI, then a byproduct might be to capture additional (historical) CINs on the MEDS cross-reference tables.

C) If the changes are relatively simple, then DHS could probably install this change this summer, but if it grows in complexity, then this will push back the schedule.

D) DHS expects that the new SAWS consortia systems will opt to use the electronic alert file (which contains raw data, including the CIN) and not use the printed alert. However, if the counties still want print, we could put their CIN somewhere in the alert header data area for both the online and printed alerts. We would just have to squeeze it in somewhere.

FEEDBACK FROM THE GROUP

Adding the ability to automatically link client data with multiple CINs is an improvement, however it is not the optimal solution. The optimal solution would be “one client, one CIN” (i.e., everyone knows the client by the same CIN).

SCI Web Page

The purpose of developing a website for the SCI application would be to provide customers with the following items:
a) Information about system problems,
b) Advance information on proposed changes
c) Access to information that would help them do a more efficient/accurate job (such as a set of best practices for file clearance)
d) A vehicle to provide input on SCI and related issues.

This site would also help other programs that are considering using SCI (such as IHSS/CMIPS, WIC, and the Statewide Immunization System) make an informed decision about whether to do so.

The content on the page would be static and would not contain any confidential or sensitive information. It would include the following information (at the most):

1. An overview of SCI
2. Trouble-shooting procedures
3. Frequently asked questions
4. Examples of how some users are making good use of the system capabilities(including best practices)
5. Information on what changes have been made within the past 60-120 days.
6. Information about what changes are planned within the next 60-120 days.
7. A list of items for which ITSD is soliciting user feedback.
8. In the event of a system failure, a description of the problem and the expected duration before it is fixed.

There is no estimated completion date for a SCI web page at this time.

System Outages

System outages are either planned or unplanned.

A planned system outage could be when DHS needs to bring the system down on a weekend so that we could run some large batch jobs. Another instance would be the upcoming Memorial Day weekend when the Health and Human Services Data Center is going to shut down most services so that they can make some changes to their operating system that will greatly reduce the risk of a power outage this year. Typically we send an e-mail note to a list of contacts in the various agencies and count on them to get the word out.

There is some degree of similarity with the notifications for unplanned outages. Marty Bornstein send notes out to the same group of folks, but obviously there is no advance notice. Most of the time, we don't know the exact duration. So Marty typically sends out a series of notes with the status until it is resolved. There is a problem, in notifying customers, if Marty happens to be out of the office when there is an outage.

FEEDBACK FROM THE GROUP

Even if SCI installs a web page, which contains information about system outages, the group still wants DHS to continue to send e-mail messages.
DHS has established a Help Desk to report, or to check on, system problems.
The phone numbers are 1-916-657-1010 or 1-800-579-0874.

Data Integrity Issues:

The accuracy of the CIN assignment directly affects the ability to link clients across programs and across time. This ability is so critical for participating State and local programs to provide services. Here is a partial list of the benefits of accurate CIN assignment:

1. Efficient and accurate sharing client data by participating programs

2. Accurate capturing client time on aid

3. Detecting fraud, duplicate aid, and instances of inappropriate dual capitation

4. Coordinating benefits and services for clients

5. Establishing effective outreach programs

6. Moving to a single point of application for DHS services and reducing the amount of redundant data requested and captured

7. Measuring health outcomes to make the optimal use of scarce public health resources

8. Increasing funding for public health programs

9. Linking the eligibility of adults on the Healthy Families Program to children on the Medi-Cal program

10. Moving towards the Administration's vision of e-government with a single portal approach.

Chances of achieving success in these various goals will be greatly enhanced if the SCI application and the customer's applications can make improvements in this area of client identification. There are two aspects to data integrity:

a) Reducing the number of new duplicates created;
b) Cleaning up old duplicates.

As more clients apply for benefit programs that require fingerprinting, this will help to reduce the number of new duplicates. Making improvements to the SCI system and to the customers systems will also help to reduce the number of new duplicates. Sharing of best practices for file clearance could help. Not all applications have incorporated the SCI duplicate check function, this would help too.

ITSD developed reports for each major user to determine the number of times they appeared to add a record to SCI for a client who was already known. The reports were the number of duplications for the 1998-2000 time period. The applications, which had the lowest number of duplicates, were SFIS and the HFP. We are working with MEB to send the findings to ISAWS and LEADER along with some recommendations. These are expected to be sent out within a few weeks.

DHS plans to modify the MEDS batch system to reduce the number of pseudo MEDS IDs issued, which will reduce the corresponding number of multiple CINs assigned to the same client.

There are a couple of issues to resolve:

a) Rules are in place for combining records in cases where both records are known to MEDS, which is a very big chunk of our records. Rules need to be developed for handling those cases where both records are not known to MEDS

b) DHS has the ability to create lists of probable duplicates. There is reluctance to have the automated system combine these records without program input. One potential exception might be in handling those cases where one of the records is a truncated MEDS record, which means that the client has not been reported to MEDS in at least 16 months and in some cases, several years. When we send out duplication findings to the various programs, we plan to send reports of actual occurrences. After we have rules established for identifying which CIN to utilize when two SCI records are combined, then we could create a new SCI transaction that customers could use.

FEEDBACK FROM THE GROUP

Although duplications are a problem, a much bigger problem is when an application links to the wrong CIN and clobbers the data on that record.

The recommendation of the group was that DHS change the way that SCI processes LINKS. If the LINK transaction were to either change multiple fields or change at least two portions of the birth date (YYYY, MM, and DD), then SCI would not apply the change, but instead return a response which tells the user that this update will change the following fields.

If the user evaluates the response and determines that this really is the client's record, then the user would submit a new SCI transaction called a FORCED LINK.

There was some discussion as to whether this capability should be limited to a group of "super-users", but the consensus was that the decision to limit access to the FORCED LINK should be up to the individual application.

There would be some reluctance to allowing an automatic combining of records without some kind of screening by a human.

Action Items

1. User Group notes will be sent out to participants.

2. DHS-ITSD will create and distribute a synopsis of the issues associated with the combining of SCI records when either only one of the records or neither of the records is known to MEDS; or when SFIS identifies that there are matching prints on file for a client.

3. DHS will convene another meeting within three months to brainstorm the issues and constraints associated with combining SCI records with the intent of providing a set of recommendations to the programs. The group recognizes that program staff in DHS and DSS must ultimately drive this process.

4. DHS-ITSD will query LEADER, SFIS, and ISAWS within the next 30-60 days to define whether there are any additional business requirements for the SCI OM transaction.

5. Volunteer counties and other programs will measure how often the client's address on the application matches the MEDS address and will be prepared to share these findings at the next user group meeting.

6. DHS will ask SSA whether the presence of data in the Middle Name field is likely to impact the percentage of SSNs that validate, either positively or negatively.

7. DHS-ITSD will perform a random sample to estimate the percentage of SCI records that could be updated with Middle Name data from DMV.

8. DHS staff will perform preliminary analysis for implementing a family-based search and will be prepared to share these findings at the next user group meeting.